Available online at www.sciencedirect.com

ScienceDirect

Information and Computation 206 (2008) 694–709

Information and Computation

www.elsevier.com/locate/ic

When unlearning helps

Ganesh Baliga [a], John Case [b,1], Wolfgang Merkle [c], Frank Stephan [d,*,2]
Rolf Wiehagen [e]

[a] *Rowan University, Glassboro, NJ 08028, USA*
[b] *University of Delaware, Newark, DE 19716-2586, USA*
[c] *Universität Heidelberg, Institut für Informatik, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany*
[d] *National University of Singapore, Departments of Mathematics and Computer Science, Singapore 117543, Republic of Singapore*
[e] *Technische Universität Kaiserslautern, Fachbereich Informatik, Postfach 3049, 67653 Kaiserslautern, Germany*

Received 5 July 2007
Available online 29 January 2008

Abstract

Overregularization seen in child language learning, for example, verb tense constructs, involves abandoning correct behaviours for incorrect ones and later reverting to correct behaviours. Quite a number of other child development phenomena also follow this U-shaped form of learning, unlearning and relearning.

A *decisive* learner does not do this and, more generally, *never* abandons an hypothesis H for an inequivalent one where it later conjectures an hypothesis equivalent to H, where equivalence means semantical or behavioural equivalence. The first main result of the present paper entails that decisiveness is a real restriction on Gold's model of explanatory (or in the limit) learning of grammars for languages from positive data. This result also solves an open problem posed in 1986 by Osherson, Stob and Weinstein.

Second-time decisive learners semantically conjecture each of their hypotheses for any language at most twice. By contrast, such learners are shown not to restrict Gold's model of learning.

Non-U-shaped learning liberalizes the requirement of decisiveness from being a restriction on *all* hypotheses output to the same restriction but only on *correct* hypotheses. The situation regarding learning power for non-U-shaped learning is a little more complex than that for decisiveness. This is explained shortly below.

Gold's original model for learning grammars from positive data, called **EX**-*learning*, requires, for success, *syntactic* convergence to a correct grammar. A slight variant, called **BC**-*learning*, requires only *semantic* convergence to a sequence of correct grammars that need not be syntactically identical to one another.

The second main result says that non-U-shaped learning does *not* restrict **EX**-learning. However, from an argument of Fulk, Jain and Osherson, non-U-shaped learning *does* restrict **BC**-learning.

In the final section is discussed the possible meaning, for cognitive science, of these results and, in this regard, indicated are some avenues worthy of future investigation.

* Corresponding author. Fax: +65 67794580.
E-mail addresses: baliga@rowan.edu (G. Baligat), case@cis.udel.edu (J. Case), merkle@math.uni-heidelberg.de (W. Merkle), fstephan@comp.nus.edu.sg (F. Stephan), wiehagen@informatik.uni-kl.de (R. Wiehagen).

1 John Case was supported in part by the National Science Foundation (NSF) Grant CCR 0208616.

2 Frank Stephan was supported by the Deutsche Forschungsgemeinschaft (DFG) Grants Am 60/9–2 and Ste 967/1–1 from 1997 to 2003 and in part by NUS Grant R252-000-212-112 from 2005 to 2007.

0890-5401/$ - see front matter
doi:10.1016/j.ic.2007.10.005

Keywords: Computational learning theory; Cognitive science; Inductive inference of grammars for languages from positive data

1. Introduction

The main topics of the present work are *decisive* and *non-U-shaped* learning which deal with constraints on returning to abandoned hypotheses. Here *decisive learning* has the constraint that whenever a learner abandons a hypothesis, it does never come back to it; *non-U-shaped learning* has the less restrictive constraint that a learner never abandons a correct hypothesis for an incorrect one and later returns to the correct one. These two criteria are parallel to *conservative learning* where the learner abandons a hypothesis only in the presence of counterexamples. Decisiveness occupied the attention of learning theorists for a number of years. In developmental and cognitive psychology, there are a number of situations in which non-U-shaped behaviour was observed; initially, this behaviour was implicitly assumed to be equivalent to decisive behaviour in inductive inference [3]. But later, a more careful analysis showed that these two criteria are different in this framework. The work presented here is the full version of a preceding conference article [3], where it has been shown that decisive learning is more restrictive than explanatory learning; in addition, the more recent result that decisiveness differs from non-U-shapedness is published here for the first time (although some follow-up papers building on these two results appeared earlier [7,8,17]).

The motivation of the whole research on decisive and non-U-shaped learning (not apparently distinguished outside the inductive inference community) stems from, for example, the following significant example discussed in empirical child cognitive development research. When studying *verb regularization in language acquisition* [22], it was observed that children first learned the correct proper forms of past tense in English language (for example, ‘called’ with ‘call’ and ‘caught’ with ‘catch’), then they overregularized and begin to form past tenses by attaching regular verb endings such as ‘ed’ to the present tense forms (even in irregular cases like ‘catch’ where that is not correct) and lastly they correctly handle the past tenses (both regular and irregular). Similar observations of U-shaped sequences for child development were made in such diverse domains as understanding of temperature [31,32], understanding of weight conservation [4,31], the interaction between understanding of object tracking and object permanence [4,31], and face recognition [5]. Within some of these domains we also see temporally separate U-shaped curves for the child’s qualitative and quantitative assessments [31,32].

One wonders if the seemingly inefficient U-shaped sequence of learning, unlearning and relearning is a mere accident of the natural evolutionary process that built us humans or something that must be that way to achieve our needed learning. We do not answer this very difficult empirical question. But, in the present paper, in the context of Gold’s formal model of language learning (from positive data) [14], we show, as a consequence of our first main theorem (Theorem 7 in Section 3 below), that there are cases where *successful* learning *requires* the learner to output hypotheses behaviourally (or semantically) equivalent to hypotheses abandoned in the learning process previously.

More precisely, as noted above, *decisive learning* [26] is learning in which the learner *cannot* conjecture an hypothesis H_1, then conjecture a behaviourally (or semantically) *inequivalent* hypothesis H_2 and then conjecture an hypothesis H_3 which *is* behaviourally equivalent to H_1 (see Definition 2 in Section 2 below). Hence, a decisive learner never semantically returns to abandoned hypotheses and, therefore, in particular, continues to output correct hypotheses from the time it has output its first correct hypothesis. Again, a consequence of our first main result in the present paper (Theorem 7) is that there are some classes of r.e. languages *learnable* from positive data which can*not* be learned decisively.

A U-shaped learner from cognitive science and as described above is formally modeled as a slight variant of a *non*-decisive learner, a variant in which, roughly, the learner semantically returns to abandoned *correct* hypotheses (but may or may not semantically return to *un*successful hypotheses). More precisely, *non-U-shaped learning* (see Definition 19 in Section 4 below) liberalizes the requirement of decisiveness from being a restriction on *all* hypotheses output to the same restriction but on correct hypotheses only.

Of course, in the light of the cognitive science motivations described above, we are extremely interested in the question as to whether non-U-shaped learning is restrictive on learning power (as decisive learning is). The answer is that it depends. To explain, we first proceed a little more formally, state some more results and then return to and answer, in a more detailed way, this question about non-U-shaped learning.

A *text* for a language L is an infinite sequence of all and only the elements of L (together with some possible #'s), that is, the elements of L might occur arbitrarily often and in any order and the #'s represent pauses. Indeed, the pause symbol is necessary as the only text for the empty language is an infinite sequence of such pauses. A text for L should be thought of as a presentation of the *positive data about* L. Gold's model of language learning from positive data [14] is also called **EX**-*learning from text*. A machine M **EX**-*learns from text* a language L iff (by definition) M, fed any text for L, outputs a sequence of grammars and this sequence eventually converges to some fixed grammar for L. Instead of type-0 grammars, one could also just use r.e. indices from a given acceptable numbering as it is often done in Recursion Theory. In a slight extension of Gold's basic model, a machine M **BC**-*learns from text* [10,27] a language L iff (by definition) M, fed any text for L, outputs a sequence of grammars such that from some point on all are grammars for L although they might not be syntactically the same. That is, **EX**-learning from text involves *syntactic* convergence to correct grammars while **BC**-learning from text involves only *semantic* or behaviourally correct convergence.

Formally, our first main result (Theorem 7), more generally says that there are classes of r.e. languages which can be **EX**-learned from text, but cannot be decisively **BC**-learned from text. From this we obtain in Corollaries 8 and 9 that decisive learning *limits* learning power for *each of* **EX**-learning and **BC**-learning from text. The latter result on **BC**-learning had been shown by Fulk et al. [12, Theorem 4]; whereas, this result on **EX**-learning is new and answers an open question of Osherson et al. [26]. In contrast, we also show (Proposition 17 in Section 4) that **EX**-learnable classes which contain the entire set of natural numbers, $\mathbb{N}$, *do* have a decisive **EX**-learner.

Note that it had been known before that, when learning programs for *functions*, decisiveness does *not* limit learning power at all (see Remark 12 for references and further explanation).

We informally define *second-time decisive learning* as learning in which, for each text input to the learner, there is *no* conjectured subsequence of hypotheses H_1, H_2, H_3, H_4, H_5 such that H_1 *is* semantically equivalent to H_3 and H_5 but semantically *in*equivalent to H_2 and H_4 (see Definition 13 in Section 4). Contrasting interestingly with our first main result, we show, in Proposition 16 in Section 4, that the learning power of second-time decisive **EX**-learners is the *same* as that of *un*restricted **EX**-learners. Hence, the additional power of non-decisive learning is already achieved if we allow the learner to "return" semantically to each abandoned hypothesis at most once.

In the next paragraph, we begin to explain what we know thus far about the power of (formal) U-shaped learning.

Another contrast to our first main result is provided by our second main result, Theorem 20 in Section 4. This result says that non-U-shaped learning suffices to learn *all* **EX**-learnable classes. Interestingly, the situation changes when **BC**-learning, rather than **EX**-learning, is considered. Indeed, we have the following remark which easily follows from a proof of Jain and Osherson [12, proof of Theorem 4].

Remark 1. Non-U-shaped **BC**-learning and unrestricted **BC**-learning do *not* coincide.

On the other hand, analogously to the case of **EX**-learning, non-U-shaped **BC**-learning and decisive **BC**-learning differ in power (see Corollary 21 in Section 4).

In brief summary, when learning languages, semantically returning to abandoned hypotheses *is* sometimes helpful for both **EX**-learning and **BC**-learning. U-shaped learning is helpful too, but for **BC**-learning only.

In Section 5, we present results showing how decisiveness interacts with a number of other restrictions on learning, for example, *conservativeness* [1], *cautiousness* [15,26], *weak-monotonicity* [18,21] and *prudence* [11,15,26] (see Definitions 22 and 24 below).

Finally, in Section 6, we more extensively summarize our principal results, provide a more detailed discussion regarding the possible power of U-shaped learning from cognitive science and indicate a number of avenues worthy of future investigation.

Our exposition of results below is interspersed with a number of additional *remarks* (besides Remark 1 above) which are useful in our proofs or of interest in their own right.

2. Decisive learning

Now, we present the definition of learning and decisiveness formally; the first basic definitions are quite general.

We use the variable σ (with or without subscripts) for finite sequences of natural numbers and the pause symbol. Such sequences are called *strings*. Initial segments of texts are always strings. The range of a string σ is the set of non-pauses in σ and is denoted by $\mathrm{rng}(\sigma)$. We write $\preceq$ for the prefix relation between strings and/or texts, for example, $\sigma_1 \preceq \sigma_2$ just in case σ_1 is a prefix of σ_2. We write $\sigma\tau$ for the concatenation of the strings σ and τ. The index $M(\sigma)$ is machine M's conjectured grammar based on the information contained in σ. $W_{M(\sigma)}$ is the language defined by the grammar $M(\sigma)$ and $W_{i,s}$, where $i, s \in \mathbb{N}$, is the set of all elements of W_i which can be enumerated within s steps of computation of the standard effective enumeration of W_i.

Definition 2. A learner M is *decisive* on a set S of strings iff there are no three strings σ_1, σ_2 and σ_3 such that σ_1 and σ_3 are in S, $\sigma_1 \preceq \sigma_2 \preceq \sigma_3$ and $W_{M(\sigma_1)}$ differs from $W_{M(\sigma_2)}$ but is equal to $W_{M(\sigma_3)}$.

A learner M is *decisive* iff it is decisive on the set of all strings.

So, a decisive learner may not *semantically* return to any output which it has previously abandoned. In particular, a decisive learner is never U-shaped and continues to output correct hypotheses from the moment it outputs its first correct hypothesis.

We conclude this section with a series of *remarks* and their proofs describing some standard techniques for the construction of decisive learners. We will apply these techniques in subsequent proofs.

Remark 3. A finite class $\boldsymbol{C}$ can always be decisively **EX**-learned.

For this result, assume that $\boldsymbol{C}$ is given as a list $L_0, L_1, \ldots, L_n$ of r.e. languages with r.e. indices $e_0, e_1, \ldots, e_n$ such that, for all i, j with $i < j \leq n$, e_i is an index for L_i and $L_i \not\supseteq L_j$. Then, for every i, j with $i < j \leq n$, there exists an element $x_{i,j}$ in $L_j - L_i$. The decisive learner uses the straightforward algorithm to output at any time the hypothesis e_i for the least i such that no $x_{i,j}$ with $j \in \{i+1, i+2, \ldots, n\}$ has been seen in the input so far.

Remark 4. If a learner M is decisive on two sets S_1 and S_2 of strings such that the classes $\{W_{M(\sigma)} : \sigma \text{ in } S_1\}$ and $\{W_{M(\sigma)} : \sigma \text{ in } S_2\}$ are disjoint, then M is actually decisive on the union of S_1 and S_2.

For a proof, assume that there were strings σ_1, σ_2 and σ_3 with $\sigma_1 \preceq \sigma_2 \preceq \sigma_3$ where σ_1 and σ_3 are strings in the union of S_1 and S_2 and $W_{M(\sigma_1)}$ is equal to $W_{M(\sigma_3)}$ but differs from $W_{M(\sigma_2)}$. In case σ_1 and σ_3 were either both in S_1 or both in S_2, M could not be decisive on the corresponding set S_i; whereas, in case one of the strings were in S_1 and the other in S_2, this would contradict the assumption on M, S_1 and S_2.

Remark 5. By delaying the learning process, we can transform a learner M_1 into a new learner M_2 for the same class such that the outputs of M_2 satisfy certain properties. Here, on input σ, the learner M_2 outputs $M_1(\gamma)$ where γ is the maximal prefix of σ such that M_2 has already been able to verify that $M_1(\gamma)$ has the property under consideration. In the remainder of this remark, we make this idea more precise and argue that, while delaying the learning process this way, we can preserve decisiveness.

Formally, we fix a binary computable predicate on strings, written in the form $P_\sigma(\tau)$, such that, for all strings σ_1, σ_2 and γ,

$$[P_{\sigma_1}(\gamma) \quad \text{and} \quad \sigma_1 \preceq \sigma_2] \qquad \text{implies} \qquad P_{\sigma_2}(\gamma) \tag{1}$$

and we define a partial function s on strings by

$$s(\sigma) = \max\{\gamma : \gamma \preceq \sigma \ \text{ and } \ P_\sigma(\gamma)\}$$

where it is to be understood that $s(\sigma)$ is defined iff the maximization in its definition is over a nonempty set. Then, by (1), the function s is nondecreasing in the sense that if s is defined on strings σ_1 and σ_2 with $\sigma_1 \preceq \sigma_2$, then $s(\sigma_1)$ is a prefix of $s(\sigma_2)$.

In case $s(\sigma)$ is defined, we let $M_2(\sigma) = M_1(s(\sigma))$ and, otherwise, we let $M_2(\sigma) = e$ for some fixed index e. We will refer to such a transformation of M_1 by the expression 'delaying with initial value e and condition P' and, informally, we will call the learner M_2, 'a delayed learner with respect to M_1.' For example, in the sequel, we will consider delayings with parameterized conditions P, first, such that $P_\sigma(\tau)$ is true iff the range of τ is contained in $W_{M_1(\tau),|\sigma|}$ and, second, such that $P_\sigma(\tau)$ is true for all σ and τ where the computation of M_1 on input τ terminates in at most $|\sigma|$ steps. The rationale for choosing these conditions will become clear in connection with the intended applications.

Now, assume that we are given a class $\boldsymbol{C}$ where, for every text T for a set in $\boldsymbol{C}$, the values of the function s have unbounded length on the prefixes of T; that is, there are arbitrarily long prefixes τ and σ of T with $\tau \preceq \sigma$ such that $P_\sigma(\tau)$ is true. Then, it is immediate from the definition of M_2 that, in case the learner M_1 learns $\boldsymbol{C}$ under the criterion **EX** or **BC**, the delayed learner M_2 learns $\boldsymbol{C}$ under the same criterion.

Finally, assume that M_1 is decisive and that either $W_e \neq W_{M_1(\sigma)}$ for all strings σ or $W_e = W_{M_1(\lambda)}$, where λ denotes the empty string. Exploiting that, in both cases, by assumption on e, as shown just below, the set $E = \{\sigma : W_{M_2(\sigma)} = W_e\}$ is closed under taking prefixes, one can show that M_2 is again decisive.

We show that E is indeed closed under prefixes, arguing by cases.

Case 1. $(\forall\sigma)[W_e \neq W_{M_1(\sigma)}]$. Suppose $\sigma \in E$. Then, by definition of M_2, $s(\sigma)\uparrow$. Let $\sigma' \prec \sigma$. If $s(\sigma')\downarrow$, we have $(\exists\gamma \preceq \sigma')[P_{\sigma'}(\gamma)]$. But, by hypothesis on P_σ, we have $P_{\sigma'}(\gamma) \wedge \sigma' \preceq \sigma \Rightarrow P_\sigma(\gamma)$. Then $s(\sigma)\downarrow$. Contradiction. Therefore $s(\sigma')\uparrow$ and $\sigma' \in E$.

Case 2. $W_e = W_{M_1(\lambda)}$, λ the empty string. Observe that, in any case, $W_e = W_{M_2(\lambda)}$ (that is, $\lambda \in E$) holds: if $P_\lambda(\lambda)$ is true then $s(\lambda)\downarrow = \lambda$ and $M_2(\lambda) = M_1(s(\lambda)) = M_1(\lambda)$ and therefore $W_{M_2(\lambda)} = W_{M_1(\lambda)} = W_e$. If $P_\lambda(\lambda)$ is false then $s(\lambda)\uparrow$ and by definition of M_2 we have $M_2(\lambda) = e$. Let $\sigma \in E$.

Case 2.1. $s(\sigma)\uparrow$. Argue as in Case 1.

Case 2.2. $s(\sigma)\downarrow$. Then, since $\sigma \in E$, we have $W_{M_1(s(\sigma))} = W_e$. Let σ' be such that $\lambda \prec \sigma' \prec \sigma$ (by the observation that $\lambda \in E$, it is sufficient to consider such σ'). If $W_{M_2(\sigma')} \neq W_e$, since by hypothesis of Case 2.2 $W_{M_2(\sigma')} = W_{M_1(s(\sigma'))}$, it follows that $W_e \neq W_{M_1(s(\sigma'))}$. Observe that by definition of the function $s(\cdot)$, for all strings τ', τ, if $\tau' \preceq \tau$, and $s(\tau')\downarrow$, then $s(\tau') \preceq s(\tau)$. But we know that $W_{M_1(\lambda)} = W_{M_1(s(\sigma))} = W_e$ and therefore $\lambda, s(\sigma'), s(\sigma)$ are a counterexample to the decisiveness of M_1.

For a proof by contradiction, assume that M_2 is not decisive, that is, there are strings σ_1, σ_2 and σ_3 with $\sigma_1 \preceq \sigma_2 \preceq \sigma_3$ such that $W_{M_2(\sigma_1)}$ is equal to $W_{M_2(\sigma_3)}$ but differs from $W_{M_2(\sigma_2)}$. If the set E intersects $\{\sigma_1, \sigma_2, \sigma_3\}$, then it must contain σ_1 by being closed under prefixes. But, then, by definition, E must contain σ_3 and, again by closure under prefixes, must contain σ_2, that is, $W_{M_2(\sigma_1)} = W_{M_2(\sigma_2)} = W_e$, thus contradicting our assumption on the σ_i's. Thus, $W_{M_2(\sigma_i)} \neq W_e$ for $i = 1, 2, 3$; hence, the strings $s(\sigma_1)$, $s(\sigma_2)$ and $s(\sigma_3)$ witness that M_1 is not decisive, giving a contradiction again.

3. The limits of decisive learning

In this section, we show that decisiveness is a proper restriction for **EX**-learning from text. In the proof of the latter result, we will use Theorem 6, which relates to a result stated by Jain et al. [15, Exercise 5–3]: every class which does not contain $\mathbb{N}$ and can be **EX**-learned, can in fact be **EX**-learned by a learner which never outputs an index for $\mathbb{N}$; that is, for all σ, $W_{M(\sigma)}$ differs from $\mathbb{N}$. Observe that Theorem 6 can be shown with **BC**-learning replaced by **EX**-learning by essentially the same proof.

Theorem 6. *Let $\boldsymbol{C}$ be an infinite class where every finite set is contained in all but finitely many sets in $\boldsymbol{C}$. If the class $\boldsymbol{C}$ can be decisively* **BC***-learned from text, then it can be decisively* **BC***-learned from text by a learner which never outputs a hypothesis for $\mathbb{N}$.*

Proof. By assumption, there is a decisive **BC**-learner M_0 which learns $\boldsymbol{C}$ from text. In case M_0 never outputs an index for $\mathbb{N}$ we are done. So, fix a string τ_0 such that $W_{M_0(\tau_0)} = \mathbb{N}$ and, by assumption on $\boldsymbol{C}$, choose $A \neq \mathbb{N}$ in $\boldsymbol{C}$ which contains $\mathrm{rng}(\tau_0)$. For every text for A, the learner M_0 must eventually output an index for A and, consequently, we can fix an extension τ of τ_0 such that $M_0(\tau)$ is an index for A. But A differs from $\mathbb{N}$ and, thus, for all extensions of τ, the decisive learner M_0 can never again output an index for $\mathbb{N}$ (hence, in particular, $\mathbb{N} \notin \boldsymbol{C}$). In the construction of a **BC**-learner M as asserted in the theorem, the key idea now is to restrict the outputs of M to indices of the form $M_0(\tau\sigma)$, except for at most finitely many additional indices of sets in $\boldsymbol{C}$ which do not contain the set $D = \mathrm{rng}(\tau)$. We partition the set of all strings into the sets

$$S_1 = \{\sigma : D \not\subseteq \mathrm{rng}(\sigma)\} \quad \text{and} \quad S_2 = \{\sigma : D \subseteq \mathrm{rng}(\sigma)\}$$

and we partition $\boldsymbol{C}$ into the classes

$$\boldsymbol{C}_1 = \{L \text{ in } \boldsymbol{C} : D \not\subseteq L\} \quad \text{and} \quad \boldsymbol{C}_2 = \{L \text{ in } \boldsymbol{C} : D \subseteq L\}.$$

By assumption on $\boldsymbol{C}$, the class $\boldsymbol{C}_1$ is finite and, as in Remark 3, we can fix a decisive **EX**-learner M_1 for $\boldsymbol{C}_1$, which learner, in particular, outputs only indices for sets in $\boldsymbol{C}_1$. Concerning $\boldsymbol{C}_2$, first consider a learner $\widetilde{M_2}$ which on input σ outputs $M_0(\tau\sigma)$. We leave to the reader the routine task of showing that $\widetilde{M_2}$ **BC**-learns $\boldsymbol{C}_2$ and inherits the property of being decisive from M_0. Let M_2 be the learner obtained, according to Remark 5, by delaying $\widetilde{M_2}$ with initial index e and condition P where e is an index for some set which neither contains D nor is in $\boldsymbol{C}_1$ and the condition P is defined by

$$P_\sigma(\gamma) \quad \text{iff} \quad D \text{ is contained in } W_{\widetilde{M_2}(\gamma),|\sigma|}.$$

We next show directly that $E = \{\sigma : W_{M_2(\sigma)} = W_e\}$ is closed under prefixes. Let $\sigma \in E$.

Case 1. $s(\sigma)\uparrow$. Argue like in Case 1 in the comment to Remark 5.

Case 2. $s(\sigma)\downarrow$. By definition we have $s(\sigma) = \max\{\gamma : \gamma \preceq \sigma \text{ and } P_\sigma(\gamma)\}$, that is, $s(\sigma) = \max\{\gamma : \gamma \preceq \sigma \text{ and } D \subseteq W_{\widetilde{M_2}(\gamma),|\sigma|}\}$. Therefore $W_{\widetilde{M_2}(s(\sigma))} \neq W_e$, since, by choice of e, $D \not\subseteq W_e$. This shows that $\sigma \in E$ iff $s(\sigma)\uparrow$, that is, that only Case 1 can hold.

Now, we obtain a learner M as required where

$$M(\sigma) = \begin{cases} M_1(\sigma) & \text{in case } \sigma \text{ is in } S_1, \\ M_2(\sigma) & \text{otherwise.} \end{cases}$$

In order to see that M **BC**-learns $\boldsymbol{C}$, assume that M is presented to a text T for a set L in $\boldsymbol{C}$. In case L is in $\boldsymbol{C}_1$, the learner M agrees on all prefixes of T with the learner M_1 for $\boldsymbol{C}_1$; while, similarly, in case L is in $\boldsymbol{C}_2$, the learner M agrees on almost all prefixes of T with the learner M_2 for $\boldsymbol{C}_2$. In order to see that M is decisive, it suffices to observe that S_1 and S_2 witness that M satisfies the assumption of Remark 4. More precisely, M_1 and M_2 are both decisive; furthermore, every set of the form $W_{M_2(\sigma)}$ is either equal to W_e or contains D, hence is distinct from every set of the form $W_{M_1(\sigma)}$. □

Theorem 7 just below is the first main result of this paper.

Theorem 7. *There is a class which can be* **EX**-*learned from text, but cannot be decisively* **BC**-*learned from text.*

From Theorem 7 the following corollaries are immediate. Here, Corollary 8 answers an open problem stated by Osherson et al. [26] and by Fulk et al. [12], while Corollary 9 has been previously shown by Fulk et al. [12]. We will argue in Remark 18 that the proof of Corollary 9 by Fulk et al. [12] neither yields Theorem 7 nor Corollary 8.

Corollary 8. *The concept of decisive* **EX**-*learning from text is a proper restriction of* **EX**-*learning from text; that is, there is a class which can be* **EX**-*learned from text, but cannot be decisively* **EX**-*learned from text.*

Corollary 9. *The concept of decisive* **BC**-*learning from text is a proper restriction of* **BC**-*learning from text; that is, there is a class which can be* **BC**-*learned from text, but cannot be decisively* **BC**-*learned from text.*

Proof of the theorem. The class $\boldsymbol{C}$ is constructed as follows. Let $M_0, M_1, \ldots$ be a recursive enumeration of all primitive recursive learners. Furthermore, define a partial-recursive function F from $\mathbb{N} \times \mathbb{N} \to \mathbb{N}^*$ such that $F(m, n)$ is the first string σ found satisfying the following conditions iff such a string exists and $F(m, n)$ is undefined otherwise:

– n is the least nonelement of the range of σ;
– $W_{M_m(\sigma)}$ is a proper superset of the range of σ.

The property of being the least nonelement is used as a type of id. That is, for any language L, $\mathrm{id}(L)$ is its first nonelement if it exists; otherwise $\mathrm{id}(\mathbb{N})$ is ∞. Furthermore, a partial K-recursive one-to-one function G with K-recursive domain is defined at input n as follows:

– If there is an $m < n/2$ such that

- $G(k) \neq m$ for all $k < n$ where $G(k)$ is defined;
- There is a $\ell \in \{m, m+1, \ldots, n\}$ such that $F(m,\ell)$ is defined and $\mathrm{id}(W_{M_m(F(m,\ell))}) = n$;

then $G(n)$ is the least such m. If no such m is found then $G(n)$ is undefined.

It can easily be verified by induction that G is K-recursive and has a K-recursive domain since the step to deal with $G(n)$ after the previous values have been processed only involves the testing whether $F(m,n)$ is defined for some $m < n/2$ and whether n is in the range of $W_{M_m(F(m,n))}$ what both can be done using the halting-problem oracle K.

Finally, these two tools, that is, the functions F and G, permit to define the class $\boldsymbol{C}$ to be learnt such that it contains for every n the following sets:

– the range of the string $\sigma = F(m,n)$ for every $m \in \{1, 2, \ldots, n\}$ where $F(m,n)$ is defined;
– the set $W_{M_{G(n)}(F(G(n),\ell))}$ for the first $\ell \in \{G(n), G(n)+1, \ldots, n\}$ with $\mathrm{id}(W_{M_{G(n)}(F(G(n),\ell))}) = n$ whenever $G(n)$ is defined;
– the set $\mathbb{N} \setminus \{n\}$ whenever $G(n)$ is undefined.

Claim 1. *$\boldsymbol{C}$ is* **EX**-*learnable.*

We construct a learner N_0 which **EX**-learns $\boldsymbol{C}$ from text. Since G is a partial K-recursive function with a K-recursive domain and since F is partial-recursive, there is a total-recursive approximation g finding an index of the following set with id n in $\boldsymbol{C}$ which is specified implicitly by G at n:

– If $G(n)$ is defined then $\lim_t g(n,t)$ converges to the index $M_{G(n)}(F(G(n),\ell))$ for the first $\ell \in \{G(n), G(n)+1, \ldots, n\}$ where $F(G(n),\ell)$ is defined and $\mathrm{id}(W_{M_{G(n)}(F(G(n),\ell))}) = n$;
– If $G(n)$ is undefined then $\lim_t g(n,t)$ converges to an index of $\mathbb{N} \setminus \{n\}$.

Given any input with range D, id n and length t, the learner N_0 conjectures an index for D if there is some $m \leq n$ such that the computation of $F(m,n)$ converges within t steps and outputs a string with range D. Otherwise N_0 conjectures $g(n,t)$. In order to enforce syntactic convergence, one can furthermore assume that N_0 outputs always the same index for the set D in the case that N_0 discovers that D equals to the range of some $F(m,n)$ with $m \leq n$.

For the verification of N_0, consider any language $L \in \boldsymbol{C}$. After N_0 has seen enough data, the following conditions hold for the parameters n, D, t on the input seen so far as above:

– the parameter n equals to $\mathrm{id}(L)$;
– g has converged at n in the sense that $g(n,t) = g(n,u)$ for all $u > t$;
– if L is finite then $D = L$;
– for all $m \in \{0, 1, \ldots, n\}$ where $F(m,n)$ is defined, $F(m,n)$ is already defined within t steps;
– for all $m \in \{0, 1, \ldots, n\}$ where $F(m,n)$ is defined and has a range not containing all elements of L, D contains at least one element outside the range of $F(m,n)$.

In the case that D is the range of $F(m,n)$ for some $m \leq n$ then N_0 remarks this as the corresponding computation has converged within t steps and outputs an index for D. In the case that D is not equal to the range of any $F(m,n)$ with $m \in \{0, 1, \ldots, n\}$, N_0 outputs $g(n,t)$ which equals to $\lim_u g(n,u)$. By construction of the approximation g, this value is the index of the only set in $\boldsymbol{C}$ besides the ranges of $F(0,n), F(1,n), \ldots, F(n,n)$ which has id n. So N_0 converges to an index of L and N_0 is an **EX**-learner for $\boldsymbol{C}$.

Claim 2. *$\boldsymbol{C}$ is not decisively behaviourally correctly learnable.*

Assume now by way of contradiction that there is a decisive **BC**-learner for $\boldsymbol{C}$. As the class $\boldsymbol{C}$ contains at most $n+1$ sets of id n for each n, $\boldsymbol{C}$ satisfies the precondition of Theorem 6 and so one can assume that there is a decisive **BC**-learner N_1 for $\boldsymbol{C}$ which never outputs any index for $\mathbb{N}$.

Now consider a delayed learner N_2 obtained by delaying N_1 with initial index $N_1(\lambda)$ and condition P where $P_\sigma(\tau)$ is true iff the computation of N_1 on input τ terminates after at most $|\sigma|$ steps. Then, N_2 is again a **BC**-learner for $\boldsymbol{C}$ which never outputs a hypothesis for $\mathbb{N}$ and, by Remark 5, is decisive. Moreover, N_2 can obviously be chosen to be primitive recursive. As a consequence, there is an index m in the listing $M_0, M_1, \ldots$ of primitive recursive learners considered above such that $M_m = N_2$.

Note that the function G is undefined at infinitely many places as G is a partial and one-to-one function mapping each k in its domain to something smaller than $k/2$. So there is a value h outside the domain of G such that $h > 2m$ and $h > k$ for all k with $G(k) < m$. As, on almost all prefixes of the ascending text for $\mathbb{N} \setminus \{h\}$, M_m outputs an index for that set, this text has a prefix σ containing $0, 1, \ldots, h-1$ such that $M_m(\sigma)$ is an index for $\mathbb{N} \setminus \{h\}$. Then σ would be a possible value for $F(m,h)$ and thus $F(m,h)$ is defined. As M_m never outputs a hypothesis for $\mathbb{N}$, the inequalities $h \leq \mathrm{id}(W_{M_m(F(m,h))}) < \infty$ hold. By the inductive construction of G, $G(\mathrm{id}(W_{M_m(F(m,h))})) = m$ unless G takes the value already at some smaller number, thus m is in the range of G.

So let n be the number with $G(n) = m$. Then there is a number $\ell \in \{m, m+1, \ldots, n\}$ such that $F(m,\ell)$ is defined and $\mathrm{id}(W_{M_m(F(m,\ell))}) = n$. Consider the two sets $L = W_{M_m(F(m,\ell))}$ and H being the range of $F(m,\ell)$. Both sets L and H are in $\boldsymbol{C}$ and $M_m(F(m,\ell))$ is an index for L. L is a proper superset of H. Furthermore there is an extension σ of $F(m,\ell)$ with range H such that $M_m(\sigma)$ is an index of H and a further extension τ of σ such that $M_m(\tau)$ is an index for L since otherwise M_m would fail to learn the sets H, L. It follows that M_m is not decisive in contradiction to the assumption. Thus $\boldsymbol{C}$ does not have a decisive behaviourally correct learner.

Theorem 7 above and the following Remark 11 show that the concepts of **EX**-learning from text and decisive **BC**-learning from text are incomparable in the sense that, for each of these concepts, there are classes which can be learned under this concept but not under the other one. In the proof in Remark 11 below, we use a standard class known to be **BC**—but not **EX**-learnable from text and show that this class can be *decisively* **BC**-learned from text.

We recall the concept of locking sequence which will be used frequently below.

Definition 10. Let a learner M, a set L and a string σ be given. Then, σ is a *locking sequence for M and L* iff $\mathrm{rng}(\sigma)$ is contained in $L = W_{M(\sigma)}$ and $M(\sigma\tau)$ is equal to $M(\sigma)$ for all strings τ over $W_{M(\sigma)}$. Furthermore, σ is a *locking sequence for M* iff σ is a locking sequence for M and $W_{M(\sigma)}$.

A learner M *learns via locking sequences* iff for every language L which is **EX**-learned by M, every text for L has a prefix which is a locking sequence for M and L.

It is known that every class which can be **EX**-learned from text at all, can actually be **EX**-learned from text via locking sequences (see, for example, Fulk [11, Theorem 13] and the references cited there).

Remark 11. The class $\boldsymbol{C}$ of all sets of the form $K \cup \{x\}$, where K is the halting problem and $x \in \mathbb{N}$, can be decisively **BC**-learned from text, but cannot be **EX**-learned from text.

For a proof, first consider a learner N where $N(\sigma)$ is an index for $K \cup \mathrm{rng}(\sigma)$. Then, N **BC**-learns $\boldsymbol{C}$ from text and, moreover, N is decisive because, for all strings τ and σ with $\tau \preceq \sigma$, the set $W_{N(\tau)}$ is contained in $W_{N(\sigma)}$.

Next assume, for a proof by contradiction, that there were a learner M which **EX**-learns $\boldsymbol{C}$ from text. Let τ be a locking sequence for M and K. Then, the set

$$H = \{x : (\exists\sigma)\,[\sigma \text{ is a string over } K \cup \{x\}, \tau \preceq \sigma \text{ and } M(\sigma) \neq M(\tau)]\,\}$$

is recursively enumerable. However, as will be shown in the next paragraph, H is the complement of K which contradicts the fact that H is recursively enumerable. Thus, $\boldsymbol{C}$ cannot have a computable **EX**-learner.

To see that H is the complement of K, note that the set H does not contain any $x \in K$ since τ is a locking sequence for M, while for $x \in K, K = K \cup \{x\}$. On the other hand, for every $x \notin K$, there is a σ as in the definition of H because, in this case, the set K is strictly contained in $K \cup \{x\}$ and, consequently, as M learns $K \cup \{x\}$, by extending τ to a text for $K \cup \{x\}$, we eventually reach a string in σ over $K \cup \{x\}$ such that σ extends τ and $M(\sigma) \neq M(\tau)$. So, every $x \notin K$ is in H.

Remark 12. Schäfer-Richter [30] (see also Osherson et al. [26, Section 4.5.5]) and, implicitly, Wiehagen [35] showed that every class of *functions* which can be **EX**-learned can in fact also be decisively **EX**-learned (notice that a basic trick from Schäfer-Richter [30] is employed, modified, in the proofs of both our Propositions 16 and 17; furthermore, note that one can think of function learning as the special case of language learning restricted to languages which are the graphs of (total) computable functions). The same holds for **BC**-learning (this is shown explicitly by Fulk et al. [12] and, implicitly, by Freivalds et al. [13]).

By contrast, again for function learning, decisiveness *does* limit learning power for the criterion $\mathbf{EX}^*$. This criterion is just like the criterion **EX** *except* that, for successful learning, the final programs can be wrong on finitely many inputs. For *this* criterion, then, for the sake of the decisiveness concept, two programs are considered *behaviourally equivalent* iff they compute the same (possibly partial) function except at finitely many inputs. Not only, then, for $\mathbf{EX}^*$, does decisiveness make a difference in learning power, but there are $\mathbf{EX}^*$ learnable classes $\boldsymbol{C}$ which can be learned only by learners which, on some $f \in \boldsymbol{C}$, must oscillate between two behaviourally *in*equivalent programs *arbitrarily finitely* many times. A particular example of such a class $\boldsymbol{C}$ is the set of all $\{0,1\}$-valued functions which are almost everywhere 0 or almost everywhere 1, that is, the set of all functions of the form σb^∞, for $b \in \{0,1\}$. Moreover, for this $\boldsymbol{C}$, for any learner of $\boldsymbol{C}$, there is a (total) function $f \notin \boldsymbol{C}$ such that, on f, the learner must oscillate *infinitely* often between programs for behavioural equivalents of functions of the two forms $\sigma 0^\infty$ and $\tau 1^\infty$.

Of course, again for the criterion $\mathbf{EX}^*$, the fact that decisiveness does make a difference in learning power extends to *language* learning: use the same classes $\boldsymbol{C}$ but construed as classes of languages which happen to be graphs of (total) computable functions.

4. The power of decisive learning

While we have shown, in the just previous section, that decisiveness properly restricts **EX**-learning, we will show now that in certain respects decisive and unrestricted **EX**-learning are rather close. We first show that every **EX**-learnable class can be learned under a criterion which is slightly more liberal than decisive **EX**-learning in so far as every abandoned hypothesis can be semantically reconjectured at most once. Then, we prove that every **EX**-learnable class can, indeed, be decisively **EX**-learned if it contains $\mathbb{N}$. Using the latter result, we finally prove our second main result (Theorem 20) that every **EX**-learnable class *can* be learned in a non-U-shaped manner. Intuitively, while a decisive learner may not semantically return to *any* abandoned hypothesis, a non-U-shaped learner is instead not allowed to return semantically to any *learnable* abandoned hypothesis (see Definition 19 below, here a hypothesis is *learnable* iff it is in the class to be learnt and consistent with all data seen so far).

Definition 13. A learner M is *second-time decisive* iff there are no five strings $\sigma_1, \ldots, \sigma_5$ with $\sigma_1 \preceq \sigma_2 \preceq \sigma_3 \preceq \sigma_4 \preceq \sigma_5$ such that $W_{M(\sigma_1)}$ is equal to $W_{M(\sigma_3)}$ and $W_{M(\sigma_5)}$ but differs from $W_{M(\sigma_2)}$ and $W_{M(\sigma_4)}$.

Remark 14. A learner M is second-time decisive if and only if there is a set S of strings such that M is decisive on S as well as on the complement of S. For a proof of this equivalence, first assume that M satisfies the right-hand side as witnessed by some set S. If there were five strings as in Definition 13 above, then, out of the three strings σ_1, σ_3 and σ_5, there would be two which either both belong to S or both belong to the complement of S, thus contradicting the fact that M is decisive on S and its complement. Conversely, if there are no five strings as in the definition, then M satisfies the right-hand side as witnessed by the set S of all strings σ_3 such that for some strings σ_1 and σ_2 with $\sigma_1 \preceq \sigma_2 \preceq \sigma_3$, the set $W_{M(\sigma_1)}$ is equal to $W_{M(\sigma_3)}$ but differs from $W_{M(\sigma_2)}$. The reason for the decisiveness on S and the complement of S is that S contains a string σ iff it has "appeared for the second time" after being abandoned intermediately while the complement of S contains a string σ iff its hypothesis had "appeared for the first time". Note that no σ is linked to a hypothesis which has "appeared for the third time or more" by assumption on M, hence every σ can be put either into S or its complement.

The subsequent proofs in this section will use Theorem 15 below. Our proof of this theorem is an adaptation of the proof of the well-known fact, noted just after Definition 10 above, that every class which can be **EX**-learned from text at all, can actually be **EX**-learned from text via locking sequences.

Theorem 15. *Let the class $\boldsymbol{C}$ be* **EX**-*learned from text by some learner M_0 and let g be a computable function such that every finite set is contained in $W_{g(n)}$ for infinitely many n.*

Then, there is a learner M and a set S of strings such that M **EX**-*learns $\boldsymbol{C}$ from text, M is decisive on S and*

$$W_{M(\sigma)} \text{ is in } \begin{cases} \{W_{M_0(\tau)} : \tau \text{ is a string}\} & \text{if } \sigma \text{ is in } S, \\ \{W_{g(i)} : i \text{ in } \mathbb{N}\} & \text{if } \sigma \text{ is not in } S. \end{cases} \tag{2}$$

Furthermore, if the sets $W_{g(0)}, W_{g(1)}, \ldots$ are mutually distinct, then M is also decisive on the complement of S.

Proof. By the comment right after Definition 10, we can assume that $\boldsymbol{C}$ is **EX**-learned by M_0 via locking sequences [26]. Observe in this connection that the standard construction for transforming a learner M_0 into a new learner that learns via locking sequences works such that all guesses of the new learner are of the form $M_0(\tau)$, that is, a conjecture of M_0 made on a string τ. Hence the theorem and, in particular, condition (2) holds for M_0 if and only if it holds for the transformed learner. Define for strings τ the following.

- τ is *consistent* iff the range of τ is contained in $W_{M_0(\tau)}$;
- τ is *s-consistent* iff the range of τ is contained in $W_{M_0(\tau),s}$;
- τ is *self-locking* iff there is no string η over $W_{M_0(\tau)}$ such that $M_0(\tau) \neq M_0(\tau\eta)$;
- τ is *s-self-locking* iff there is no string η over $W_{M_0(\tau),s}$ of length at most s such that $M_0(\tau) \neq M_0(\tau\eta)$.

By definition, a string τ is a locking sequence for M_0 if and only if τ is consistent and self-locking. Furthermore, a string is consistent if and only if it is s-consistent for almost all s, and a string is self-locking if and only if it is s-self-locking for all s.

Fix a one-to-one computable function i that maps pairs of a string and a natural number to natural numbers such that for all τ and s, the set $W_{g(i(\tau,s))}$ contains the range of τ as well as $W_{M_0(\tau),s}$. Then, let, for each string τ,

$$V(\tau) \quad = \quad \begin{cases} W_{M_0(\tau)} & \text{if } \tau \text{ is self-locking,} \\ W_{g(i(\tau,s))} & \text{if } s > 0 \text{ is minimal such that } \tau \text{ is not } s\text{-self-locking.} \end{cases}$$

Claim 1. *The sets $V(\tau)$ are recursively enumerable and an index $v(\tau)$ for $V(\tau)$ can be computed from τ.*

A Turing machine which enumerates the set $V(\tau)$ might, for example, for stages $s = 0, 1, 2, \ldots$, enumerate the numbers in $W_{M_0(\tau),s}$, while, in case it encounters a stage s such that τ is not s-self-locking, the Turing machine enumerates the numbers in the superset $W_{g(i(\tau,s))}$ of $W_{M_0(\tau),s}$ and stops. The construction of this Turing machine is effective in τ, hence the function v can be chosen to be computable. This completes the proof of the claim.

For the remainder of this proof, say σ is *connected* to σ' if $\sigma \preceq \sigma'$ and

$$M_0(\sigma) = M_0(\gamma) \text{ for all } \gamma \text{ with } \sigma \preceq \gamma \preceq \sigma'.$$

Define a partial function τ from strings to strings as follows. Let

$\tau(\sigma)$be the shortest $\tau \preceq \sigma$ such that
- τ and σ are connected,
- τ is $|\sigma|$ – self-locking,
- τ is $|\sigma|$ – consistent.

By construction, the function τ is partial-recursive and its domain is a recursive set. As a consequence, the following learner M is indeed computable.

$$M(\sigma) = \begin{cases} V(\tau(\sigma)) & \text{if } \tau(\sigma) \text{ is defined,} \\ g(i(\sigma, 0)) & \text{otherwise.} \end{cases}$$

By substituting the definition of V, we obtain

$$W_{M(\sigma)} = \begin{cases} W_{M_0(\tau(\sigma))} & \text{if } \tau(\sigma) \text{ is defined and self-locking,} \\ W_{g(i(\tau(\sigma),s))} & \text{if } \tau(\sigma) \text{ is defined and not self-locking,} \\ W_{g(i(\sigma,0))} & \text{if } \tau(\sigma) \text{ is undefined,} \end{cases} \tag{3}$$

where $s > 0$ is the value from the definition of $V(\tau(\sigma))$. We show that the assertion of the theorem is satisfied for M and the set

$$S = \{\sigma\colon \tau(\sigma) \text{ is defined and self-locking}\}.$$

From the definition of S and by (3) it is immediate that condition (2) in the theorem is satisfied.

Claim 2. *The learner M* **EX**-*learns* $\boldsymbol{C}$ *from text.*

Given a text T for a language L in $\boldsymbol{C}$, by assumption on M_0, there is a least prefix τ_0 of T which is a locking sequence for M_0 and L. That is, $M_0(\tau_0)$ is an index for L and τ_0 is consistent and self-locking. By choice of τ_0, for almost all prefixes σ of T the string τ_0 but no proper prefix of τ_0 satisfies all three conditions in the definition of $\tau(\sigma)$. Hence for almost all prefixes σ of T, we have $\tau(\sigma) = \tau_0$ and $M(\sigma) = v(\tau_0)$, where $v(\tau_0)$ is an index for $L = W_{M_0(\tau_0)}$ by (3) and because τ_0 is self-locking.

Claim 3. *Let τ be defined on strings σ and σ' where $\sigma \preceq \sigma'$. Then $\tau(\sigma) \preceq \tau(\sigma')$; that is, the function τ is nondecreasing.*

By definition of τ, the strings $\tau(\sigma)$ and $\tau(\sigma')$ are connected prefixes of σ and σ', respectively. In particular, both strings are prefixes of σ', hence the claim follows if one can show that $\tau(\sigma')$ is not a proper prefix of $\tau(\sigma)$. For a proof by contradiction, assume that the latter is false, that is, we have $\tau(\sigma') \prec \tau(\sigma) \preceq \sigma \preceq \sigma'$. In case σ is not connected to σ', we obtain as a contradiction that $\tau(\sigma')$ but not its extension $\sigma \preceq \sigma'$ is connected to σ'. In case σ is connected to σ', the string $\tau(\sigma')$ is connected to σ because it is connected to σ', is $|\sigma|$-self-locking because it is $|\sigma'|$-self-locking where $|\sigma'| \geq |\sigma|$ and is $|\sigma|$-consistent because its extension $\tau(\sigma)$ is $|\sigma|$-consistent. That is, $\tau(\sigma')$ satisfies all three conditions in the definition of τ at place σ, thus contradicting the minimality of $\tau(\sigma)$.

Claim 4. *If $\tau(\sigma)$ is defined and self-locking, then $\tau(\sigma')$ is defined and coincides with $\tau(\sigma)$ for all σ' to which σ is connected.*

Fix σ as in the assumption of the claim and some σ' to which σ is connected. Then $\sigma \preceq \sigma'$, hence $\tau(\sigma) \preceq \tau(\sigma')$ by Claim 3. On the other hand, $\tau(\sigma') \preceq \tau(\sigma)$ by the minimization in the definition of τ. More precisely, $\tau(\sigma)$ is connected to σ' because it is connected to σ, is $|\sigma'|$-self-locking because it is self-locking and is $|\sigma'|$-consistent because it is $|\sigma|$-consistent.

Claim 5. *Let σ_1, σ_2 and σ_3 be strings where $\sigma_1 \preceq \sigma_2 \preceq \sigma_3$ and $\tau(\sigma_1)$ and $\tau(\sigma_3)$ are defined and coincide. Then $\tau(\sigma_2)$ is defined and coincides with $\tau(\sigma_1)$ and $\tau(\sigma_3)$.*

By Claim 3, it suffices to show that $\tau(\sigma_2)$ is defined at all, that is, that there is a string that satisfies all three conditions in the definition of τ at place σ_2. The latter is witnessed by the string $\tau(\sigma_1) = \tau(\sigma_3)$, which is connected to σ_2, is $|\sigma_2|$-self-locking, and $|\sigma_2|$-consistent because it is connected to both of σ_1 and σ_3, is $|\sigma_3|$-self-locking and is $|\sigma_1|$-consistent, respectively.

Claim 6. *The learner M is decisive on S.*

Fix any strings σ_1 and σ_3 in S and a string σ_2 such that $\sigma_1 \preceq \sigma_2 \preceq \sigma_3$; recall that by definition of S, the strings $\tau(\sigma_1)$ and $\tau(\sigma_3)$ are consistent and self-locking. In case σ_1 is connected to σ_3, by Claims 4 and 5, the function τ has the same value on σ_1, σ_2, and σ_3. In case σ_1 is not connected to σ_3, then also $\tau(\sigma_1)$ is not connected to $\tau(\sigma_3)$, that is, there is a string γ such that $\tau(\sigma_1) \preceq \gamma \preceq \tau(\sigma_3)$ and $M_0(\tau(\sigma_1))$ differs from $M_0(\gamma)$. But $\tau(\sigma_1)$ is self-locking, thus there must be a number which is not contained in $W_{M_0(\tau(\sigma_1))}$ but is contained in the range of γ. Hence this number is contained in the range of $\tau(\sigma_3)$ and its superset $W_{M_0(\tau(\sigma_3))}$. Consequently, the sets $W_{M(\sigma_1)} = W_{M_0(\tau(\sigma_1))}$ and

$W_{M(\sigma_3)} = W_{M_0(\tau(\sigma_3))}$ are distinct. In summary, the claim follows because for any strings σ_1, σ_2 and σ_3 as above either the sets $W_{M(\sigma_i)}$ for $i = 1, 2, 3$ are all the same or $W_{M(\sigma_1)}$ differs from $W_{M(\sigma_3)}$.

Claim 7. *If the sets $W_{g(0)}, W_{g(1)}, \ldots$ are mutually distinct, then M is decisive on the complement of S.*

For any distinct pairs (γ, s) and (γ', s'), the set $W_{g(i(\gamma,s))}$ differs from $W_{g(i(\gamma',s'))}$ because i is one-to-one and by assumption on the sets $W_{g(i)}$. Furthermore, by (3), for any string σ not in S, in case $\tau(\sigma)$ is undefined, the set $W_{M(\sigma)}$ is equal to $W_{g(i(\sigma,0))}$, whereas in case $\tau(\sigma)$ is defined, $W_{M(\sigma)}$ is equal to $W_{g(i(\tau(\sigma),s))}$ for some $s > 0$ that only depends on $\tau(\sigma)$. So, given any σ_1 and σ_3 not in S such that $W_{M(\sigma_1)}$ is equal to $W_{M(\sigma_3)}$, the strings $\tau(\sigma_1)$ and $\tau(\sigma_3)$ are both defined and are the same. Then, by Claim 5, for any σ_2 such that $\sigma_1 \preceq \sigma_2 \preceq \sigma_3$, the value of $\tau(\sigma_2)$ is defined and coincides with $\tau(\sigma_1)$, hence $W_{M(\sigma_1)} = W_{M(\sigma_2)}$. The claim now follows by definition of decisiveness. Furthermore, with this claim, the proof of Theorem 15 is completed.

We have seen in Section 3 that there are classes which can be **EX**-learned from text, but cannot be decisively **EX**-learned. Proposition 16 shows that the additional power of non-decisive learning is already achieved if we allow the learner to return semantically to each abandoned hypothesis at most once.

Proposition 16. *Every class which can be* **EX**-*learned from text can also be* **EX**-*learned from text by a second-time decisive learner.*

Proof. By Remark 14, Proposition 16 is a special case of Theorem 15 where we fix a computable function g such that, for all i, the set $W_{g(i)}$ is just $\{0, \ldots, i\}$. $\square$

The class constructed in the proof of Theorem 7 in order to separate the concepts of unrestricted and decisive **EX**-learning does not contain the set $\mathbb{N}$. By Proposition 17 just below, which is again a direct consequence of Theorem 15, this is no coincidence.

Proposition 17. *Every class which contains $\mathbb{N}$ and can be* **EX**-*learned from text can be decisively* **EX**-*learned from text.*

Proof. Let $\boldsymbol{C}$ be **EX**-learnable and contain $\mathbb{N}$. Fulk [11] showed that $\boldsymbol{C}$ has a prudent learner M_0 – such a learner outputs only indices of sets which it also learns. Since there is a locking sequence τ for $\mathbb{N}$, M_0 does not learn any finite language containing $\mathrm{rng}(\tau)$. Hence, defining $W_{g(i)} = \{0, 1, \ldots, i + \max \mathrm{rng}(\tau)\}$ makes the sets $W_{g(i)}$ different from all sets learned by M_0 and, thus, also different from all sets conjectured by M_0.

We apply Theorem 15 to M_0 and g. We obtain an **EX**-learner M for $\boldsymbol{C}$ and a set S of strings such that M is decisive on both S and its complement. Moreover, $W_{M(\sigma)}$ and $W_{M(\eta)}$ are different for all $\sigma \in S$ and $\eta \notin S$; hence, M is already decisive by Remark 4. $\square$

Remark 18. Fulk et al. [12, Theorem 4] give an example of a class $\boldsymbol{L}$ which can be **BC**-learned from text but cannot be learned so decisively. While their construction bears some similarities to the construction of the class $\boldsymbol{C}$ in the proof of Theorem 7, their class $\boldsymbol{L}$ is not **EX**-learnable from text and, consequently, neither yields Theorem 7 nor a separation of decisive and unrestricted **EX**-learning as stated in Corollary 8.

For a proof that $\boldsymbol{L}$ is not **EX**-learnable, note that there is no set in $\boldsymbol{L}$ which contains the numbers $\langle 0, 0\rangle$ and $\langle 1, 0\rangle$, where $\langle \cdot, \cdot \rangle$ denotes a standard pairing function [29]. Hence, by switching to some fixed index for $\mathbb{N}$ as soon as the data contain both numbers, every **EX**-learner for $\boldsymbol{L}$ can be transformed into an **EX**-learner for $\boldsymbol{L}$ which also identifies $\mathbb{N}$. But, then, by Proposition 17, if the class $\boldsymbol{L}$ were **EX**-learnable, it were decisively **EX**-learnable; whereas, by construction, $\boldsymbol{L}$ is not even decisively **BC**-learnable.

Proposition 17 above gives a sufficient condition for an **EX**-learnable language class to be also decisively learnable. This condition can be both easily described and checked. Clearly, if we drop this condition, then, by Corollary 8, we can no longer guarantee decisive learning. But, surprisingly, we *can* guarantee non-U-shaped learning (for **EX**), yet the latter is, nonetheless, seemingly close to decisive learning.

Definition 19. A learner M is *non-U-shaped* on a class $\boldsymbol{C}$ of languages iff for any $L \in \boldsymbol{C}$ and any text T for L, there are no prefixes σ_1, σ_2, and σ_3 of T such that $\sigma_1 \preceq \sigma_2 \preceq \sigma_3$, $W_{M(\sigma_1)}$ and $W_{M(\sigma_3)}$ are equal to L but differ from $W_{M(\sigma_2)}$.

Thus, non-U-shaped learning liberalizes the requirement of decisiveness from being a restriction on *all* hypotheses output to the same restriction but on *learnable* outputs only. Our second main result follows.

Theorem 20. *Every* **EX**-*learnable class is non-U-shaped* **EX**-*learnable.*

Proof. Let $\mathcal{C}$ be any **EX**-learnable class and let M_0 be a corresponding learner. We may assume that $\mathbb{N}$ is not in $\mathcal{C}$, since, otherwise, by Proposition 17, $\mathcal{C}$ would be decisively learnable and, hence, non-U-shaped learnable. Let e_0 be an index for $\mathbb{N}$ and let g be the constant function with value e_0. Apply Theorem 15 to g and M_0 in order to obtain a set S and an **EX**-learner M for $\mathcal{C}$, where M is decisive on S, while $W_{M(\sigma)}$ is equal to $\mathbb{N}$ for all strings σ that are not in S. But $\mathcal{C}$ does not contain $\mathbb{N}$, hence M is non-U-shaped on $\mathcal{C}$. □

Note that Theorem 20 is no longer correct if we consider **BC**-learning in place of **EX**-learning. Actually, non-U-shaped **BC**-learning *differs* from unrestricted **BC**-learning (see Remark 1). On the other hand, analogously to **EX**-learning, non-U-shaped **BC**-learning and decisive **BC**-learning are different as well.

Corollary 21. *The concept of decisive* **BC**-*learning from text is a proper restriction of non-U-shaped* **BC**-*learning from text; that is, there is a class which can be non-U-shaped* **BC**-*learned from text, but cannot be decisively* **BC**-*learned from text.*

Proof. Of course, decisive learning implies non-U-shaped learning.

Let $\mathcal{C}$ be the class from the proof of Theorem 7. By this proof, $\mathcal{C}$ is not decisively **BC**-learnable. On the other hand, $\mathcal{C}$ is **EX**-learnable and, hence, by Theorem 20, $\mathcal{C}$ is non-U-shaped **EX**-learnable as well. But, then, obviously, $\mathcal{C}$ is non-U-shaped **BC**-learnable. □

5. Decisive, conservative, cautious and prudent learning

In cases where we already know that some concept of learning does not allow the learning of all **EX**-learnable classes, by using Proposition 17, we can often show that the concept under consideration in fact does not even allow the learning of all classes which can be decisively **EX**-learned. We show this now for the concepts of conservative and cautious **EX**-learning defined just below (Definition 22).

Definition 22. A learner M is *conservative* iff for all strings σ_1, σ_2 with $\sigma_1 \preceq \sigma_2$ and $\mathrm{rng}(\sigma_2) \subseteq \mathrm{W}_{\mathrm{M}(\sigma_1)}$, the hypothesis $M(\sigma_2)$ is equal to $M(\sigma_1)$ [1].

A learner M is *cautious* iff for all strings σ_1, σ_2 with $\sigma_1 \preceq \sigma_2$, the set $W_{M(\sigma_2)}$ is not a proper subset of $W_{M(\sigma_1)}$ [26].

Angluin has shown that in the context of text learning, conservative **EX**-learning is a proper restriction of **EX**-learning (see Angluin [1, Theorem 4] or Jain et al. [15, Proposition 5.46]). It can further be shown that conservative **EX**-learning is a proper restriction of cautious **EX**-learning, which in turn is a proper restriction of **EX**-learning (see Exercises 4.5.4A and 4.5.4B in Osherson et al. [26]). By the latter remark, Proposition 23 just below trivially remains true with 'cautiously' replaced by 'conservatively' — and this is just the content of Exercise 4.5.5.C in Osherson et al. [26].

Proposition 23. *There is a class which can be decisively* **EX**-*learned from text, but cannot be cautiously* **BC**-*learned from text.*

Proof. Let $\mathcal{C}$ contain the set $\{x, x+1, \ldots\}$ if $x \notin K$ and all finite sets with minimum x if $x \in K$. A decisive learner can be obtained as follows: Let an input with nonempty range D and length t be given, let $x = \min(D)$; then the decisive learner conjectures the set $\{x, x+1, \ldots\}$ if $x \notin K_t$ and the set D if $x \in K_t$. But given any recursive **BC**-learner M for this class, the set of all x such that M overgeneralizes on some input with x being the minimum of the input's range is r.e. and contains $\overline{K}$. Thus it must be a proper superset of $\overline{K}$ and there is a finite set D with minimum in K on which M overgeneralizes. As M learns this set, M is not cautious. □

While we have just seen that the concepts of conservative and cautious **EX**-learning differ from decisive **EX**-learning, Proposition 25 below shows that all three concepts coincide for prudent learners.

Definition 24. A learner M is *prudent* iff for all strings σ, M learns the set $W_{M(\sigma)}$ [26].

A learner M is *weak-monotonic* iff for all strings σ and τ with $\mathrm{rng}(\sigma\tau) \subseteq \mathrm{W}_{\mathrm{M}(\sigma)}$ it holds that $W_{M(\sigma)} \subseteq W_{M(\sigma\tau)}$ [18].

The concept of weak-monotonic learner was introduced by Lange and Zeugmann [21]. In their paper, it is shown that a class of uniformly recursive languages can be conservatively **EX**-learned from text if and only if it can be **EX**-learned from text by a weak-monotonic learner. Then, the same result was proved but for arbitrary classes of r.e. languages (see Jain and Sharma [16] and Kinber and Stephan [19]).

Proposition 25. *Let the class* $\mathcal{C}$ *be* **EX**-*learned from text by some learner* M *which is decisive and prudent. Then,* M *is weak-monotonic and, in particular,* $\mathcal{C}$ *can be* **EX**-*learned from text by a conservative learner.*

Proof. Assume for a proof by contradiction that M is not weak-monotonic. Then, there are strings σ_1 and σ_2 with $\sigma_1 \preceq \sigma_2$ such that $\mathrm{rng}(\sigma_2)$ is contained in $W_{M(\sigma_1)}$ but $W_{M(\sigma_1)}$ is not contained in $W_{M(\sigma_2)}$; hence, in particular, the two latter sets are distinct. Now, M is prudent and, hence, **EX**-learns $W_{M(\sigma_1)}$. Therefore, by extending σ_2 to a text T for $W_{M(\sigma_1)}$, we obtain an extension σ_3 of σ_2 where the sets $W_{M(\sigma_1)}$ and $W_{M(\sigma_3)}$ are the same. But we have already seen that $W_{M(\sigma_1)}$ differs from $W_{M(\sigma_2)}$; hence, contrary to our assumption, M is not decisive. □

6. Summary, discussion and further problems

Corollary 26 just below summarizes the most important of the results of the present paper. Not included in Corollary 26 is the interesting separation (for language learning from text) between **EX***-learning and decisive **EX***-learning mentioned at the end of Remark 12 and the relations between decisive, conservative, cautious and prudent learning given in Section 5.

Corollary 26. *In the following diagram, an inclusion holds iff it is shown by an arrow or can be derived by following several arrows.*

$$
\begin{tikzcd}
 & \textbf{BC} & \\
 & \textbf{Non U-Shaped BC} \arrow[u] & \\
\begin{array}{c}\textbf{Non U-Shaped EX}\\ = \textbf{EX} = \\ \textbf{Second Time Decisive EX}\end{array} \arrow[ur] & & \textbf{Decisive BC} \arrow[ul] \\
 & \textbf{Decisive EX} \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

Proof. The equalities follow from Proposition 16 and Theorem 20. The inclusions are, then, straightforward. The non-inclusions follow from Remarks 1 and 11 and Theorem 7 (together with several inclusions and equalities). □

We can conclude from Corollary 26 that, for learning languages from text, semantically returning to abandoned hypotheses *is* helpful for both **EX**-learning and **BC**-learning. However, between these two criteria, U-shaped learning is helpful for **BC**-learning only. In particular, then, semantically returning to abandoned hypotheses is helpful for *both* criteria if and only if the abandoned hypotheses are *incorrect*. In this way, the present paper contains new results and insights compared to the conference version [3]. Hence, it is able to present a more nearly complete and accurate picture of decisive and non-U-shaped learning. Based on this, subsequent work investigated the role of non-U-shapedness and decisiveness with respect to other learning criteria [6,7,17]. Furthermore, Carlucci et al. [8] investigated variants of the U-shaped paradigm like the criterion which forbids to return to abandoned incorrect conjectures while one can abandon and reconsider the correct conjecture many times. Next we discuss the possible connection between the formal results obtained in the present work and the cognitive science motivation for studying (the formal version of) U-shaped learning (mentioned in Section 1 above).

The studies concerning U-shaped learning in human cognitive development that are cited above in Section 1 do not (and, perhaps, could not) give us some idea as to whether *every* human child, say, above a certain level of mental ability, *always* shows U-shaped learning in *at least one* domain. We would expect to see this *if* U-shaped learning is *necessary* to learn the things humans learn and that were somehow important in the past for surviving long enough to leave offspring who do the same. Perhaps there is at least one bright human child who seemingly exhibits *no* U-shaped learning in any known domain. If so, we might conclude that U-shaped learning is not a *necessary* phenomenon (for humans). So, what light might the formal results above shed on the problem? For example, is human learning more like **EX**-learning or **BC**-learning? Gold [14], for example, argues from the empirical psycholinguistic literature, for **EX**-learning from text; however, **BC**-learning ostensibly had not yet been considered at the time of his paper. Indeed, Gold's model of **EX**-learning from text has been extensively discussed as to what it may say about human linguistic learning—see, for example, [20,28,34]. Osherson and Weinstein [25] and Wexler [33] discuss some positive and negative aspects of modeling human linguistic learning by **BC**-learning instead of **EX**-learning. Case [9] discusses, in this regard, some other, related and intermediate criteria of learning. Again in the context of human linguistic learning, McNeill [23] notes empirically that there is *faster* learning of language for children in homes in which more corrections (usually in the form of expansions) are given. These corrections provide, among other things, *negative* information—which is not available from mere text. Baliga et al. [2] present a formal result to the effect that the presence of minimal negative (non-text) information can yield a significant improvement in language learning *speed* (as calibrated by number of mind-changes required to reach a correct grammar). It seems fair, then, to say that it is not yet clear which, if any, known variants of **EX**-learning might better model even human linguistic learning.

In general, nth time decisive and non-n-U-shaped learning (with the obvious definitions), for the many criteria other than **EX** (such as **BC**) and wholy (or partly) involving language learning from text and which are from [2, 9,10,27] have not been fully investigated herein and it could be quite interesting to do so in the future. Referring to these further criteria: the story on non-U-shaped learning and variants vis a vis their *possible* relation to cognitive development may become more complex.

Furthermore, it would be very interesting to obtain *characterizations* (especially those insightful for cognitive science) for n-th time decisive and non-n-U-shaped learning when they differ from the underlying, unrestricted criteria.

Acknowledgments

We are very grateful to Lorenzo Carlucci and the anonymous referees for a number of useful remarks and corrections.

References

[1] Dana Angulin, Inductive inference of formal languages from positive data, Information and Control 45 (1980) 117–135.
[2] Ganesh Baliga, John Case, Sanjay Jain, Language learning with some negative information, Journal of Computer and System Sciences 51 (1995) 273–285.
[3] Ganesh Baliga, John Case, Wolfgang Merkle, Frank Stephan, Unlearning helps, in: U. Montanari, J. Rolim, E. Welzl, Proceedings of the 27th International Colloquium on Automata, Languages and Programming, LNCS, vol. 1853, Springer, 2000, pp. 835–844.
[4] T. Bower, Concepts of development, in: Proceedings of the 21st International Congress of Psychology, Presses Universitaires de France, 1978.
[5] Susan Carey, An analysis of a learning paradigm, in: S. Strauss, R. Stavy (Eds.), U-Shaped Behavioral Growth, Developmental Psychology Series, Academic Press, 1982.

[6] Lorenzo Carlucci, John Case, Sanjay Jain, Frank Stephan, Non U-shaped vacillatory and team learning, in: Algorithmic Learning Theory: Sixteenth International Conference, ALT 2005, Singapore, October 2005, Proceedings, LNAI, vol. 3734, Springer, 2005, pp. 241–255.
[7] Lorenzo Carlucci, John Case, Sanjay Jain, Frank Stephan, Memory-limited U-shaped learning, in: G. Lugosi, H. Simon (Eds.), Proceedings of the 19th Annual Conference on Computational Learning Theory (COLT 2006), LNAI, vol. 4005, Springer, 2006, pp. 244–258.
[8] Lorenzo Carlucci, Sanjay Jain, Efim Kinber, Frank Stephan, Variations on U-shaped learning, Information and Computation 204 (2006) 1264–1294.
[9] John Case, The power of vacillation in language learning, SIAM Journal on Computing 28 (1999) 1941–1969.
[10] John Case, Chris Lynes, Machine inductive inference and language identification, in: M. Nielsen, E.M. Schmidt, Proceedings of the 9th International Colloquium on Automata, Languages and Programming, LNCS, vol. 140, Springer, 1982, pp. 107–115.
[11] Mark Fulk, Prudence and other conditions on formal language learning, Information and Computation 85 (1990) 1–11.
[12] Mark Fulk, Sanjay Jain, Daniel Osherson, Open problems in “Systems That Learn”, Journal of Computer and System Sciences 49 (1994) 589–604.
[13] Rusins Freivalds, Efim Kinber, Rolf Wiehagen, On the power of inductive inference from good examples, Theoretical Computer Science 110 (1993) 131–144.
[14] E. Mark Gold, Language identification in the limit, Information and Control 10 (1967) 447–474.
[15] Sanjay Jain, Daniel Osherson, James Royer, Arun Sharma, Systems that Learn: An Introduction to Learning Theory, MIT Press, 1999 (second ed. of Ref. [26]).
[16] Sanjay Jain, Arun Sharma, Generalization and specialization strategies for learning r.e. languages, Annals of Mathematics and Artificial Intelligence 23 (1998) 1–26.
[17] Sanjay Jain, Frank Stephan, Some recent results in U-shaped learning, in: Theory and Applications of Models of Computation: Third International Conference, TAMC 2006, Beijing, China, May 15–20, 2006, Proceedings, LNCS, vol. 3959, Springer, 2006, pp. 421–431.
[18] Klaus-Peter Jantke, Monotonic and non-monotonic inductive inference, New Generation Computing 8 (1991) 349–360.
[19] Efim Kinber, Frank Stephan, Language learning from texts: mind changes, limited memory and monotonicity, Information and Computation 123 (1995) 224–241.
[20] David Kirsh, PDP learnability and innate knowledge of language, in: S. Davis (Ed.), Connectionism: Theory and Practice, Oxford University Press, 1991, pp. 297–322.
[21] Steffen Lange, Thomas Zeugmann, Monotonic versus non-monotonic language learning, in: Proceedings of the Second International Workshop on Nonmonotonic and Inductive Logic, LNAI, vol. 659, Springer, 1993, pp. 256–269.
[22] Gary Marcus, Steven Pinker, Michael Ullman, Michelle Hollander, T. John Rosen, Fei Xu, Overregularization in Language Acquisition, Monographs of the Society for Research in Child Development, vol. 57, no. 4, University of Chicago Press, 1992.
[23] David McNeill, Developmental psycholinguistics, in: F. Smith, G.A. Miller (Eds.), The Genesis of Language, MIT Press, 1966, pp. 15–84.
[24] Piergiorgio Odifreddi, Classical Recursion Theory, North Holland, Amsterdam, 1989.
[25] Daniel Osherson, Scott Weinstein, A note on formal learning theory, Cognition 11 (1982) 77–88.
[26] Daniel Osherson, Michael Stob, Scott Weinstein, Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists, MIT Press, 1986.
[27] Daniel Osherson, Scott Weinstein, Criteria of language learning, Information and Control 52 (1982) 123–138.
[28] Steven Pinker, Formal models of language learning, Cognition 7 (1979) 217–283.
[29] Hartley Rogers, Theory of Recursive Functions and Effective Computability, McGraw-Hill, New York, 1967. Reprinted from: MIT Press, 1987.
[30] Gisela Schäfer-Richter, Über Eingabeabhängigkeit und Komplexität von Inferenzstrategien, Ph.D. Thesis, Rheinisch-Westfälische Technische Hochschule Aachen, Germany, 1984.
[31] Strauss Sidney, Ruth Stavy (Eds.), U-Shaped Behavioral Growth, Developmental Psychology Series, Academic Press, 1982.
[32] Sidney Strauss, Ruth Stavy, N. Orpaz, The child’s development of the concept of temperature, unpublished manuscript, Tel-Aviv University, 1977.
[33] Kenneth Wexler, On extensional learnability, Cognition 11 (1982) 89–95.
[34] Kenneth Wexler, Peter W. Culicover, Formal Principles of Language Acquisition, MIT Press, 1980.
[35] Rolf Wiehagen, Characterization problems in the theory of inductive inference, in: G. Ausiello, C. Böhm, Proceedings of the Fifth International Colloquium on Automata, Languages and Programming, LNCS, vol. 62, Springer, 1978, pp. 494–508.